Evaluation of ERA-20cm reanalysis dataset over South Korea

By

Dong-Ik Kim* and Dawei Han

Water and Environment Research Group, Department of Civil Engineering,

University of Bristol, Bristol, UK

* Corresponding Author: Queen's Building, University Walk, Clifton, Bristol, UK, BS8 1TR (E-mail : dikim0127@gmail.com)

Abstract

Long term climate data are key in assessing water related hazards in order to adapt and mitigate climate change. Reanalysis has been developed as a surrogate for local observations, but there is a lack of studies about the suitability at different parts of the world. In this study, our primary goal of this study was to identify the applicability of the ECMWF 20th century atmospheric model ensemble (ERA-20cm) in South Korea. Thus, we have evaluated the ensemble for precipitation and temperature by assessing the correlation coefficients, the long-term trend by the Mann-Kendall test, the skill score based on the probability density functions (PDFs) and the goodness of the ensemble spread. The relationship between the spread and the El Niño-Southern Oscillation (ENSO) has also been explored. ERA-20cm ensemble has difficulty in providing useful information on the long term trend as well as the temporal variability in South Korea, but, for the pdf-based comparison, all ensemble predictions represent significant agreements. It is found that the ensemble mean can misrepresent ten individual members, especially for statistical estimates, in regional-scale analyses. The ensemble does not spread well enough to cover the observation and there is no relationship between the spread and ENSO. This paper shows that the applicability of ERA-20cm may vary by region, hence these findings help to fill in the knowledge gaps about the applicability of the reanalysis in regional scale study including South Korea.

Keywords : ERA-20cm, interanuual variability, non-parametric trend test, skill score, reanalysis

1. Introduction

To adapt and mitigate climate change, it is essential to analyse the reliable long-term climate dataset. It is generally considered that the gauged local data provide the best accuracy in the gauging points, but they are usually sparse and limited in the time range (Becker et al., 2013; Simmons et al., 2004). For this reason, the reliable gridded dataset, called "reanalysis", derived using modern data assimilation techniques has been considered as a surrogate for local observations since 1990s. Representatively, the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Oceanic and Atmospheric Administration (NOAA) have produced these kinds of products such as ERA-40 from 1957 to 2002, NCEP/NCAR Reanalysis from 1948 to present, ERA-interim from 1979 to present and NCEP-Department of Energy (DOE) reanalysis from 1979 to present (Dee et al., 2011; Kalnay et al., 1996; Kanamitsu et al., 2002; Uppala et al., 2005). These half-century reanalysis datasets were produced by assimilating a wide range of available observations for the reference time (Compo et al., 2011). However, the local gauged data prior to mid-twentieth century were mainly observed over land surface in specific regions and sparse, so there was vulnerability of generating reliable products for the early 20th century in these models (Compo et al., 2011; Donat et al., 2016). Nevertheless, a few century-long reanalyses such as the ECMWF 20th century atmospheric model ensemble (ERA-20cm) and NOAA-CIRES 20th century reanalysis v2c(20CR) have been developed by the advanced techniques which generate ensembles to account for uncertainties by the relatively sparse input data (Compo et al., 2011; Hersbach et al., 2015; Poli et al., 2016; Poli et al., 2013). These ensemble predictions are produced by applying different initial conditions and observations in their own atmospheric models. For instance, for 20CR, spanning 1850 to 2014 with the 1.875°×1.9° resolution, only surface pressure observations were assimilated with an Ensemble Kalman Filter, and monthly sea surface temperature(SST) and sea-ice cover(SIC) were used as boundary conditions (Compo et al., 2011; Donat et al., 2016). In contrast, ERA-20cm, covering 1900-

2010, was produced with the Integrated Forecasting System (IFS) version Cy38r1, but it was simulated by observational SST and SIC with no data assimilation (Hersbach et al., 2015; Poli et al., 2016). Due to the difference of these simulation conditions, each ensemble model provides different representations of the observations (Gao et al., 2016). Hence, despite the availability of these reanalyses, their qualities are still an important issue in the climate change study.

In order to examine the quality of reanalysis datasets, there have been a lot of global-, continental-, or local-scale studies, and the evaluation of the ensemble predictions also has been explored (Donat et al., 2016; Donat et al., 2014; Ferguson and Villarini, 2012; Gao et al., 2016; Simmons et al., 2004). For example, Simmons et al. (2004) evaluated ERA-40 and NCEP/NCAR reanalysis by comparing them with CRU, the interpolated global gridded observations, for air temperature at 5°×5° resolution on global and continental scales and concluded that there were very similar interannual patterns between ERA-40 and CRU, especially in Northern Hemisphere from 1979 onward. In another global study, Donat et al. (2016) compared 20CR, ERA-20c and ERA-20cm with the interpolated gridded observations (HadEX2), and suggested that these reanalyses agreed well after about 1950 although they often had discrepancies during the early twentieth century. In the case of national scale evaluation, Ferguson and Villarini (2012) Compared 20CR over the central United States with CRU and suggested that there were inhomogeneities for 20CR from 1940 to 1950 during the warm season so it was recommended to use the second half century of it, not all period, over the central United States. A recent study over China by Gao et al. (2016) statistically evaluated ERA-20cm. After comparing the ensemble ten members at 0.5°×0.5° grids for precipitation and temperature, it was concluded that generally all ensemble simulations were able to represent the real condition on a comparable level.

It is important that comparative studies should cover a wide range of locations around the world and gaps should be filled in for the sites lacking such studies so that a clear pattern could be understood. In Korea, the long-term climate trend analysis on precipitation and temperature has generally been based on the observed values and the time range of these studies were limited in the late 20th century

(Bae et al., 2008; Chang and Kwon, 2007; Chung et al., 2004; Chung and Yoon, 2000; Jung et al., 2011). There were a few trials to apply the reanalysis products on the climate trend research over Korea, but these datasets were applied to estimate the features of the comparable region like East-Asia as a whole, not Korea itself at the country level (Choi et al., 2016; Ho et al., 2003; Jeong et al., 2015). In other words, the climate datasets were used in Asian area in order to compare with the climate trend of Korea examined by the observation data. However, if researchers would like to extend the analysis period up to the early of the 20th century, it is essential to attempt to find out the reliable long-term dataset with high resolution, which should be explored.

Of the 20th century ensemble reanalyses, ERA-20cm is one of the representative datasets. While 20CR provides the ensemble mean and the spread with the relatively coarse resolution, 1.875°×1.9°, in the public web server, ERA-20cm is able to support all predictions of ten ensemble members at a higher resolution such as 0.5°×0.5°. Thus, in this study, we focus on whether ERA-20cm can provide reliable data in regional-scale analyses over South Korea. It is known that ERA-20cm cannot reproduce actual synoptic situation, but it is able to detect the long term trend of the climate variables such as temperature as well as providing statistically meaningful values (Hersbach et al., 2015). However, this perception is mainly based on the ensemble mean in global- or continental-scale analysis (Hersbach et al., 2015; Poli et al., 2016), and the evaluation based on the regional-scale is relatively rare. Gao et al. (2016) assessed ERA-20cm ensemble members over China, but it was limited in statistical analyses. Kim and Han (2018) evaluated the ERA-20cm mean compared with other century-long datasets in South Korea, but did not analyse individual ensemble members. For this reason, we have evaluated ERA-20cm ensemble members for precipitation and temperature, which are commonly used in hydro-meteorological analysis, over South Korea in this study. By estimating the correlation coefficient r, the significance of trend by the Mann-Kendall test, the skill score based on the probability density functions (PDFs) and the percent of the observation within the ensemble spread, this paper assesses the temporal variability and statistical agreement of each ensemble member, and the goodness of the ensemble spread in South Korea. The relationship between the spread and the El Niño-Southern Oscillation (ENSO) has also been explored. To show the specific process, the data and methodology applied in this study are introduced in Section 2 and Section 3. Section 4 presents the main results for precipitation and temperature and finally the discussion and conclusions are described in Section 5 and Section 6, respectively.

2. Data

2.1. Observed Local Data

To analyse the precipitation and temperature change over the mainland of South Korea, daily total precipitations and daily mean 2-m air temperatures of 13 ground gauge stations are taken from the data archive of Korea Meteorological Administration(KMA) (https://data.kma.go.kr/cmmn/main.do) and merged to the monthly values. The stations are evenly selected excluding islands of Korea from 1961 to 2010 with no empty values, although three of stations are available from 1966, 1968 and 1973, separately. The quality of the observations is strictly controlled by KMA. The detailed information of the location and data period of the stations is given in Fig. 1 and Table 1.

[Insert Fig. 1]

[Insert Table 1]

2.2. ERA-20cm

The ECMWF provides ERA-20cm data with 10-member ensemble from January 1900 to December

2010 (Hersbach et al., 2015). This dataset was produced with the Integrated Forecasting System (IFS) version Cy38r1. In the system, sea-surface temperature (SST) and sea ice cover (SIC) products provided by the Met Office Hadley Center, called HadlSST2, are prescribed. HadISST2 consists of an ensemble of ten realizations (Hersbach et al., 2015). The ten realizations are a random drawing from a large SST ensemble. They are all equally likely, so the selection of allocation has no effect on the statistical feature of ERA-20cm. Here, variation in these realizations account for uncertainties in the available observational sources, affecting the status of ERA-20cm members. ERA-20cm includes no data assimilation (Donat et al., 2016). In order to find out the general feature of the ERA-20cm ensemble, we extract ten different kinds of ensemble members for 3-hourly total precipitation and 2-m air temperature with 0.5°×0.5° grid from January 1901 to December 2010 via the ECMWF server. The products in South Korea are accumulated into monthly data, and the values on the sea are excluded as in previous studies, since the comparable long-term weather stations are available in land areas and the ocean grid points would be heterogeneous to the land points for averaging (Donat et al., 2016; Hersbach et al., 2015; Kim and Han, 2018; Poli et al., 2016). To evaluate differences between ensemble predictions and the mean, ten ensemble datasets and the mean values of them representing the average of ten ensemble numbers per day in a given grid are applied in the analysis, separately. In the paper, the 10 different kinds of ensemble values will be abbreviated as from En0 to En9 hereafter, while the means are referred to as Mean.

3. Methodology

3.1. Evaluation of interannual variability

To explore the temporal variability of each ensemble member compared with the observed values, the Pearson's linear correlation coefficients(r) between the ensemble and the observations of 13 stations from 1961 to 2010 are calculated. For this analysis, the seasonal/yearly total precipitation and mean temperature variables are derived from all the dataset. Every seasonal dataset is collected for spring from March to May, summer from June to August, autumn from September to November, and winter from December to February.

For the calculation of r in a given station, we have interpolated the data in each station point by the popular inversed distance method (Babak and Deutsch, 2009). Although a grid point provides the information for a cell with 0.5°×0.5° (approx. 55km), it is clear that there is an abrupt change for the information such as averaged elevation in the boundary. Considering some stations are almost located in the middle of grids, despite the error caused by the interpolation approach, this study has applied the aforementioned method as follows :

$$\boldsymbol{w(x,y)} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \boldsymbol{\alpha_i w_i}\,, \quad \boldsymbol{\alpha_i} = \frac{\left(\frac{\boldsymbol{1}}{\boldsymbol{d_i}}\right)^{\boldsymbol{p}}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \left(\frac{\boldsymbol{1}}{\boldsymbol{d_i}}\right)^{\boldsymbol{p}}} \tag{1}$$

where N is the number of the grids used in calculation, w is the evaluated value from the data product in each station point, w_i is the i-th data point among the selected values, d_i is the distance from the station to the i-th grid, and p is the specified weighting power. In this analysis, the closest 9 gridded values from each station are used and the most common value, 2, for the power p is applied. After calculating the r values in 13 stations, the mean of them is compared.

The r values between ensemble members for annual variables are evaluated in the same way to identify the independence of ten ensemble members. In order to find out the relationship between the El Niño-Southern Oscillation (ENSO) and the ensemble spread of ERA-20cm, this study also assesses the r values between 3 month moving ensemble spread and the Oceanic Niño Index (ONI), which is the standard that NOAA applies for detecting El Niño and La Niña events. In the ONI, when three month running means of SST anomalies in the Niño 3.4 region (5°N-5°S, 120°-170°W) exceed a threshold of ±0.5°C for at least 5 consecutive months, the first month of them is considered as the start of El Niño/ La Niña.

3.2. Trend test

To find out the significance of trends in each dataset, the Mann-Kendall test is applied for the reference period 1961 – 2010. Additionally, the trends of ERA-20cm mean and ensemble members from 1901 to 2010 are compared to evaluate the significance of trends for over 100 years. The Mann-Kendall trend test created by Mann (1945) and Kendall (1955) is one of the widely used nonparametric tests for detecting linear and non-linear trend of environmental data such as precipitation, temperature and streamflow (Bae et al., 2008; Shadmani et al., 2012; Xu et al., 2005). In the Mann-Kendall test, the significance of trends is evaluated by comparing the standardised test statistic Z with the standard normal variate at the desired significance (Hamed and Rao, 1998). When $|Z| > Z_{1-\alpha/2}$, where $Z_{1-\alpha/2}$ is the standard normal deviates where the significance level is α, the null hypothesis is rejected and it means that there is a significant trend in the time series in the test. In this study, 0.05 and 0.10 are applied for the significance level. Because this method does not quantify the magnitude of the trend, the slope of the trend is estimated by Theil-Sen approach (Sen, 1968; Theil, 1950), defined by the median value of the ranked slope estimates as follows :

$$\boldsymbol{\beta} = \boldsymbol{Median}\left(\frac{\boldsymbol{x_j} - \boldsymbol{x_i}}{\boldsymbol{j} - \boldsymbol{i}}\right), \quad \forall \boldsymbol{i} < \boldsymbol{j} \tag{6}$$

In this equation, the positive value of β represents the increasing trend over time, while the negative value means the opposite trend. The advantage of this method is that it is less sensitive to outliers or extreme values than the least-square method (Sayemuzzaman and Jha, 2014; Shadmani et al., 2012).

3.3. PDF-based Evaluation method

To assess the statistical similarity between the observations and each dataset from 1961 to 2010, we have estimated the skill score based on the probability density function (PDF) suggested by Perkins et al. (2007). This method is a useful measure for comparison between two PDFs. By calculating the overlapped area between two distributions at each bin, this skill estimates how much the climate dataset distribution is similar to the observed. If a dataset matches the observed value perfectly in PDF, the skill score will be 1, which equals the sum of the probability. Otherwise, if the skill score is close to zero, it means that there is no common area between the model values and observations. In other words, the more overlapped the two curves, the closer to 1 this score is. The skill score is calculated as follows:

$$S_{score} = \sum_{1}^{n} minimum(P_m, P_0), \tag{7}$$

where n is the number of bins for the calculation, P_m is the frequency of values in a given bin from a comparison target, and P_0 is the frequency values in a given bin from observations. In this study, the square root of 1mm month^{-1} for precipitation and 1°C for temperature are considered as the intervals of bins for the monthly dataset analysis to effectively compare the PDFs like earlier studies (Gao et al., 2016; Perkins et al., 2007). The advantages of this evaluation method are that it shows more credible climate variations than the traditional mean-based method and it is flexible to collect data with different time periods from multiple stations (Gao et al., 2016; Perkins et al., 2007).

3.4. Criterion for the goodness of ERA-20cm ensemble

In this study, the goodness of ensemble spread from 1961 to 2010 is evaluated by the percent of observations within the ensemble interval, which is identical to the P-95CI method used by Li et al. (2010). This index is simply calculated as follows :

$$P(\%) = \frac{N_{in}}{n} \times 100 \tag{8}$$

where n is the number of bins and N_{in} is the number of observations bracketed by the upper and lower boundary values. In this analysis, the closeness of the P to 100% judges the goodness of the predictions. In Li et al. (2010), the 95% confidence interval was derived from thousands of simulations. However, because ERA-20cm has only ten ensemble predictions, we use the lowest and highest values at a given bin as the lower and the upper boundaries (i.e. ensemble interval).

4. Results

4.1. Precipitation

4.1.1. Interannual variability

Table 2 quantitatively explains the seasonal or annual correlation between the observation and the simulated precipitation from 1961 to 2010. In the seasonal comparison, the *r* values for all ten ensemble members are located between -0.171 and 0.228, and Mean has the values between -0.035 and 0.278. This indicates that there are little temporal correlations between ERA-20cm ensemble members as well as Mean and the observations for precipitation. The annual comparison also suggests the similar result. The highest value is 0.158 for En6, the lowest one is -0.065 for En1 and Mean performs 0.103. In other words, the *r* values of ERA-20cm ensemble predictions and Mean are overall close to zero.

[Insert Table 2]

[Insert Fig. 2]

Fig. 2 which illustrates the seasonal and annual precipitation change of each dataset from 1961 to 2010 supports this result. From the seasonal changes, it is difficult to observe the similarity between Mean or each member and the observation as well as between each ensemble predictions in any season (Fig. 2(a)). Moreover, the simulations for spring and winter have generally higher rainfalls than the observations, while summer rainfalls for ERA-20cm are underestimated. In terms of the magnitude of interannual variability, Mean fluctuates less than any other datasets including the observations. The annual change is similar to the seasonal trend (Fig. 2(b)). However, due to the aggregation effect of all seasonal values, the gap between ERA-20cm and the observation is mitigated, although there is still no correlation between them.

[Insert Table 3]

In order to identify the relationship among ten ensemble members, the correlation coefficients between ensemble member predictions are assessed on the basis of annual total precipitation analysis (Table 3). The r values are between -0.113 and 0.338, which implies that each ensemble member of ERA-20cm has little correlation with others. I.e., it is shown that the ensemble members of ERA-20cm for precipitation over South Korea are generally independent.

[Insert Fig. 3]

Fig. 3 describes the relationship between the ensemble spread of ERA-20cm and ENSO. For El Niños, there were strong events in 1972-73, 1982-83, and 1997-98, as seen in Fig. 3. However, it is difficult to find out ensemble variance which has patterns correlating with El Niños. Likewise, there were strong La-Niñas in 1973-1976, 1988-89 and 1998-2001, but there is no significant agreement between the two graphs. The r value between 3 month moving average spread of ERA-20cm and the ONI, 0.052, confirms that there is no correlation between them. In other words, ERA-20cm ensemble spread is independent to ENSO over South Korea.

4.1.2. Long-term trend

Table 4 shows the long-term trends derived by the Mann-Kendall test. The standardised statistics (Z) for the reference period 1961 to 2010 describe that there are no significant seasonal/annual trends in ensemble members and Mean except a few cases, whereas observation has the significant increasing trend in summer at 95% confidence level. Although the slopes (β) of ensemble members indicate the positive or the opposite values, it does not mean that there are significant trends. Of the members, only En2 in autumn and En6 in summer show the significant increasing trends during the study period.

[Insert Table 4]

The longer term analysis on ERA-20cm from 1901 to 2010 shows the relatively different result. Mean indicates the significant upward trends in spring, winter and annul tests, but there is no ensemble member which has the same pattern as Mean. Between the ensemble members, it is difficult to find out the similarity. En3 and En5 have the significant increasing trends in spring at 95% confidence level, while En2 and En6 have them in autumn. En4 has the positive trend in both spring and autumn at 90% confidence level. In summer, En3 and En9 indicate the significant opposite trends each other. En7 shows the obviously increasing trend in winter simulation at 95% confidence level, whereas En9 is the only ensemble with an upward trend in annual test, except for Mean. On the other hand, En0, En1 and En8 have no significant trends in all simulations. The trend difference between the datasets implies that the ERA-20cm reanalysis data have difficulty in representing the real temporal variability.

4.1.3. Statistical comparability

Despite the aforementioned inconsistency of temporal change between each ensemble member and the observation, Fig. 4 describes that there are statistically significant agreements between the observation and reanalysis data from 1961 to 2010. The skill scores of all ten ensemble members exceed 0.8, and the minimum and maximum values are 0.81 for En6 and 0.86 for En2, separately. On the other hand, the mean of ERA-20cm indicates 0.67 which is less than any ensemble predictions.

[Insert Fig. 4]

The specific reason of low skill score for Mean as well as the general characteristics of each dataset is shown in Fig. 5(a) which illustrates the PDFs of the observation, ensemble members and Mean over South Korea from 1961 to 2010 and Fig. 5(b) which represents seasonally subdivided PDFs. It is obvious that ERA-20cm reanalysis slightly underestimates the probability of dry months and intensive rainfall months and overestimates the moderate months in Fig. 5(a). Especially, Mean has the obviously overestimated probability for the moderate intensity, which comes from the temporally smoothed values in Fig. 2. Seasonally, spring and winter rainfalls for ensemble members are generally overestimated and the PDFs of summer rainfalls for ten members are located in the left of the PDF of the observation (Fig. 5(b)). Comparing with the ensemble predictions, Mean has the more exaggerated moderate values in every season. This evaluation suggests that Mean for ERA-20cm needs a cautious approach to use in the frequency analysis, albeit it as well as all ensemble members shows the significant agreement with the observation.

[Insert Fig. 5]

4.1.4. The goodness of ERA-20cm ensembles

In order to evaluate the suitability of the ensemble spread for ERA-20cm, the percent of the observations bracketed by the ensemble intervals has been calculated. Table 5 represents the result of this analysis. In terms of the temporal change spread, the values perform the moderate to high goodness with the percentage between 56 and 74. This result implies that the ensemble for precipitation is quite well spread to reproduce the real world, but it still needs the improvement to cover the whole observation curve for the reference period 1961-2010 as illustrated in Fig. 2. For the statistical analysis, the P value, 42.3, seems to indicate the moderate goodness but it plummets to 34.6 excluding the on-boundary values. This result indicates that the coverage of the ensemble is not statistically well spread, although each ensemble dataset has the significant agreement with the observation.

[Insert Table 5]

4.2. Temperature

4.2.1. Interannual variability

Table 6 describes the *r* values between the gauged temperature and the model temperature from 1961 to 2010. In seasonal simulations, ERA-20cm ten ensemble members have generally low to moderate values with 0.048 to 0.525, whereas Mean performs moderate to high correlation with the highest values between 0.471 and 0.691 in every season. To be more specific, spring and summer generally have the highest values except En4, while winter has the weakest correlation in each dataset. In annual comparison, Mean performs the significant correlation with the observation (0.746), whereas all ten members have the moderate correlations ($0.443 < r < 0.604$).

[Insert Table 6]

[Insert Fig. 6]

Fig. 6 details the seasonal and annual mean temperature patterns for each dataset in South Korea from 1961 to 2010. For the seasonal comparisons, the movements of ensemble members in a given season have partial similarity with the observation (Fig. 6(a)). In terms of values themselves, each member has generally lower temperatures than the real in all seasons except winter. The mean values for ensemble members are about 1 to 2 Celsius degrees lower than the observations from spring to autumn, and even in winter, gaps are observed since 1987 (Fig. 6(a)). From this seasonal feature, it is clearly shown that the annual reanalysis has the cooler temperature than the annual mean observed temperature, albeit there are some partial correlations between them (Fig. 6(b)).

[Insert Table 7]

To determine the independence among the ten members, the *r* value between ensemble predictions is estimated based on annual mean temperature analysis from 1961 to 2010 (Table 7). The coefficients are generally low to moderately high with the values between 0.268 and 0.682. The most correlated members are En6 and En9 with 0.682, while the least ones are En2 and En4. This analysis suggests that for temperature, there is partial covarying relationship between ensemble members as well as between the ensemble and the observation.

[Insert Fig. 7]

Fig. 7 illustrates the relationship between the ensemble spread for temperature and ENSO. As aforementioned in the comparison for precipitation, the strong El Niño happened in 1972-73, 1982-83, and 1997-98 and La-Niña happened in 1973-1976, 1988-89 and 1998-2001 (Fig. 7). Comparing this with the ensemble spread, there is little correlation between them. The r value is -0.006 between 3 month moving average spread of ERA-20cm and the ONI, which also supports this result. I.e, the ensemble spread of ERA-20cm for temperature is not affected by ENSO over Korea.

4.2.2. Long-term trend

Table 8 describes the seasonal and annual patterns for mean temperature by the Mann-Kendall approach. The seasonal analysis from 1961 to 2010 suggests that the observation and the ensemble predictions have different trends. In winter, the observation has a most strongly increasing trend at 0.05 significance level, but for ensemble members, there are no significant trends except En0 and En4. In contrasts, in summer, there is no significant trend for the observation, but ensemble members except En3 have the obvious upward trends. Likewise, Mean has the increasing trend at 95% confidence level in summer, but there is no significant trend in winter. In terms of annual analysis, each member as well as the observation and Mean shows the significant upward trend at 95% confidence. However, the slopes of annual trends for Mean and the members are generally less than the observation's. This result implies that the discrepancy between ERA-20cm and the observation for temperature, shown in Fig. 6(b), has been widened during the study period.

[Insert Table 8]

Unlike the first analysis, the simulation for 20th century indicates that all datasets have the obvious increasing trends at 0.05 significance level in every season. To be more specific about the seasonal intensities (β), the values for spring and summer generally are higher than that for winter, although all the slopes are between 0.44 and 1.39 °C per 100 year. In case of the annual simulation, it is clear that all simulations show the significant upward trends at 95% confidence level, and the extent of increase for ensembles are close to 1°C per 100 year, which is slightly smaller than those of the first simulation from 1961 to 2010. This analysis implies that in the long term, the temperature over South Korea is on the rise, although ERA-20cm may have different trends with actual climate change.

4.2.3. Statistical comparability

Fig. 8 represents the skill score of the PDF of each member for monthly mean temperature from 1961 to 2010. The estimates of all ensemble predictions are between 0.69 and 0.75, while Mean represents 0.59 which is the lowest value among the results. It shows that each dataset has the meaningful agreement with the observation, but still needs the significant improvement, especially for Mean.

[Insert Fig. 8]

Fig. 9 which illustrates the PDFs of the observed and the modelled dataset for temperature supports the skill score analysis. In Fig. 9(a), the performances of the ensemble members show high agreements in the range of below 0°C. However, there is a clear discrepancy in the range of about 0°C to 15°C, and over 15°C, the PDFs for the members are left-biased comparing with that for the observation. The seasonal comparisons show more clearly the points of the discrepancies. In Fig. 9(b), comparing the peaks of each PDF, the seasonal distributions for all ten members are generally located in the left side of the observations' except winter. This result reconfirms that ERA-20cm ensemble represents the statistically significant agreement with the observation, but it needs some bias correction before using them in the frequency analysis for temperature. Moreover, the magnitude of the peaks for Mean is higher than those of other predictions, especially in spring and autumn, which comes from the temporally averaged values in Fig. 6. It means that Mean can narrowly interpret the seasonal or annual temperature variability of reanalysis data in frequency analysis over Korea. For reference, the peaks in Fig. 9(b) are due to the rapid change in monthly mean temperature.

[Insert Fig. 9]

4.2.4. The goodness of ERA-20cm ensembles

Table 9 represents the simulation results on the basis of the ensemble interval for temperature. For interannual comparison, the seasonal/annual P(%) values represent low performances except winter. In other words, the ERA-20cm ensemble for temperature in South Korea cannot cover the historical climate change enough, as shown in Fig. 6 representing the lower temperatures of the ensemble than the observations'. For statistical test, the value is 37.1%, but it goes down to 28.6% excluding the on-boundary values. This result indicates that the spread for ERA-20cm ensemble is too narrow and biased to represent Korea's observation, although each ensemble dataset has the good skill score.

[Insert Table 9]

5. Discussion

As aforementioned in Introduction, it has been known that despite the inconsistency in synoptic events, ERA-20cm can represent the long term as well as statistical estimates of climate variables in global- or continental-scale analyses (Hersbach et al., 2015; Poli et al., 2016). However, it is also found that there still exists knowledge gap about the applicability of ERA-20cm in regional-scale studies. In this context, we have evaluated the long-term change and statistical estimate of ERA-20cm ensemble for precipitation and temperature over South Korea as a substitute for reliable long term in-situ data.

In the temporal variability comparison for precipitation, the r values for ERA-20cm mean and ensemble members compared with the observation derived from the 13 gauged stations are close to 0. This result reconfirms the aforementioned feature of ERA-20cm which cannot reproduce the real synoptic situation for precipitation in Korea. For the trend test on precipitation, there is no significant trend except En6 and the observation in summer and En2 in autumn for the reference period 1961 to 2010. However, the simulation from 1901 to 2010 shows that each ensemble member has its own trend. For instance, Mean has the increasing trends in spring, winter and 12-month average simulations, but En0, En1 and En8 have no trends. In summer, only En9 has an increasing pattern at 95% confidence level, while En3 indicates the opposite trend. En3, En4 and En5 have the significant upward trends in spring, whereas En2, En4 and En6 have the same trends in autumn. It is clear that the result of the trend analysis can vary depending on the study period and regions in South Korea (Bae et al., 2008). Nevertheless, the previous long term trend researches have shown that summer precipitation observed in Korea has generally increased (Chang and Kwon, 2007; Choi et al., 2009; Jung et al., 2011; Wang et al., 2006). Chang and Kwon (2007) and Jung et al. (2011) suggested that all stations had increasing summer rainfalls since 1973. Choi et al. (2009) compared the gauged rainfalls of 10 Asian countries from 1955 to 2007 and described the significant increasing summer rainfall in South Korea at 95% confidence level. The longest trend analysis on Seoul, the capital of South Korea, also indicated a significant upward trend from 1778 to 2004, although the estimate for the pre-1950 period suggested no significant trend (Wang et al., 2006). In the same vein, the weak trends in summer for ensemble members and Mean imply that the data are able to distort information on the climate change patterns over South Korea. I.e, it may not be a proper reanalysis to detect the long term trend for precipitation, although the previous study on a global scale showed a fair agreement with long term temperature trend (Hersbach et al., 2015).

For statistical evaluation for precipitation, there are significant agreements between the monthly averaged observations derived from 13 gauged stations and each dataset. The skill scores for all ten ensemble members exceed 0.8 as the same as Gao et al. (2016), which concluded that despite the spatial variability, all ten ensemble members of ERA-20cm for precipitation had the high skill scores, over 0.8, in China. An interesting point is that Mean has the lowest value (0.67) because it exaggerates the moderate intensity than any other members. In other words, the temporally averaged precipitation can show the smoothed movement in Fig. 2, but it results in distorting the statistical weight of the moderate intensity (Fig. 4 and 5). In some comparative studies, the mean of ERA-20cm represented the ERA-20cm ensemble itself (Donat et al., 2016; Poli et al., 2016). However, the findings in our study show that the mean is able to underestimate the dry and intensive rainfall season in Korea. Hence, it could be concluded that the mean of ERA-20cm ensemble should be carefully used in the regional study.

In terms of ensemble itself for precipitation, there are no temporal correlations between ensemble members. According to Hersbach et al. (2015), the ten members were designed to represent the different realizations and to account for uncertainties based on the random plausible SST ensemble. From the same vein, the result in this study reconfirms the independence of ensemble members. On the other hand, the P values, representing the goodness of the spread, show significant goodness for temporal variability, but statistically have the low agreement with the observation. Considering the purpose of the ensemble, which is to consider uncertainties in model data (Poli et al., 2016), this result shows that the spread has difficulty in representing the real rainfall in Korea. In other words, it needs to be widened in its coverage. In the comparison between ensemble spread and the ONI, the result shows that the spread for precipitation has no correlation with El Niño or La Niña. Reminding that ERA-20cm was generated by using SST as the boundary condition, it can be hypothesised that the relationship may influence on the ERA-20cm spread. However, unlike the previous studies which verified the significant relationships between the rainfall patterns in Korea and El Niño/La Niña (Jin et al., 2005; Son et al., 2014), the result indicates no correlation between the spread and ENSO. Jin et al. (2005) suggested that La Niña phenomena correlates with the monthly rainfall with a lag time of 4 or 5 months in Korea. Son et al. (2014) addressed that there was a significant positive correlation between El Niño and winter precipitation and especially, it was strong in early winter.

For temperature, ERA-20cm mean has moderate correlations with the seasonal/annual observations, whereas ten members have low to moderate correlations. Comparing with the result for precipitation, this result seems to show a significant relationship between the ensemble and the observed. However, the temperature trends from 1961 to 2010 indicate that the ensemble predictions have different trends from the observation. In the analysis, ERA-20cm ensemble members as well as Mean show significant increasing movements in spring and summer except a few cases, while there are no clear trends in winter excluding En0 and En4. On the other hand, the observation has a significant increasing trend in winter, and there is no significant trend in summer. According to the previous observed trends in South Korea for the late 20th century, it was concluded that the winter and annual mean temperature had the significant upward trends but the summer trend was weak (Choi et al., 2009; Chung and Yoon, 2000; Jung et al., 2002). Comparing this conclusion with the result in this paper, it could be deducted that ERA-20cm which shows the strong summer and weak winter trends has little reliability in the seasonal trend test for the late 20th century. An interesting point is that the second trend assessment from 1901 to 2010 indicates the significant warming trends in all the simulations at the 0.95 confidence level. This trend has also been shown in recent researches. Donat et al. (2016) suggested the warming trends over the world in their multi data sources analysis from 1901 to 2010, and Harris et al. (2014) showed the annual warming trend in East Asia, 0.11°C/decade, by using CRU from 1901 to 2008. However, the relatively low intensity of the slopes in winter still suggests that it still needs the cautious approach.

In the case of PDFs analysis, all ensemble members for temperature have the significant agreements to the observation with the values between 0.69 and 0.75. On the other hand, Mean has the lowest value, 0.59, because the temporally averaged temperature (Fig. 6) overestimates the range of moderate values in the pdf analysis (Fig. 8 and 9). This simulation indicates that ERA-20cm ensemble has the significant reliability for the monthly frequency for temperature in Korea, but still it is challenging to apply them, especially ERA-20cm mean, directly. In terms of regional applicability, the evaluation on ERA-20cm over China, Gao et al. (2016) showed that the skill scores of all ten ensembles exceeded 0.9 for temperature averaged over all regions. However, the scores in this study indicate the much lower values than them. This implies that applicability may vary by region, and this should be explored before using the dataset in the regional scale study.

For the ensemble itself, the *r* values between ensemble members show the partial correlations with the values between 0.268 and 0.682, although they were generated from the random SST ensemble (Hersbach et al., 2015). On the other hand, the goodness test for the spread performs the low agreement in both temporal and statistical comparisons. This shows that the spread for temperature is too narrow to cover the actual data. In terms of the relationship between the spread and ENSO, this paper clearly suggests no correlation between them, although the recent research (Lee and Julien, 2016) suggested that the teleconnection between ENSO and the temperature was clearly identified over South Korea. This demonstrates that El Niño and La Niño did not affect the ensemble spread of ERA-20cm for temperature as well as precipitation.

Due to the improved atmospheric model technique, it is logical to hypothesise that the latest reanalysis dataset should be sufficiently accurate in terms of temporal and statistical variability. However, the results in this study imply that the perception to ERA-20cm could be wrong. Previous studies have also documented that reanalysis datasets have their own uncertainty which vary in space and time (Bao and Zhang, 2013; Bosilovich et al., 2008; Kim and Han, 2018; Ma et al., 2009). There may exist various reasons for the differences. First of all, the inhomogeneity of available input data may affect the uncertainty in the modelled value (Poli et al., 2016; Thorne and Vose, 2010). I.e., as the number of available data vary in time and space, the uncertainty in ERA-20cm may be locally variable. The significant inconsistencies in extreme climate events could be one of the causes (Befort et al., 2016; Donat et al., 2016; Krueger et al., 2013). Befort et al. (2016) indicated that the century-long reanalysis like ERA-20c had difficulty in representing low-frequent extra-tropical cyclones and windstorms. Donat et al. (2016) documented that for heavy precipitation, the local agreement of century-long reanalyses was spatially variable while the global time series performed well. Regional climate regime could also be one of the reasons. Shah and Mishra (2014) documented that the reanalysis like ERA-20cm had the errors for precipitation and temperature in monsoon season over India. The elevation difference between the model and actual terrain could affect the result in mountainous regions including South Korea (Gao et al., 2012; Gao et al., 2014a; Gao et al., 2014b; Zhao and Fu, 2006). The scale difference between weather stations (point scale) and grid points (area average) may influence on the biases, especially for the extreme (Maraun, 2013). As the gridded data can be smoothed, the comparison with the weather station could be biased.

Likewise, the century-long reanalysis, ERA-20cm, necessarily includes the biases for a variety of reasons. Thus, to prevent misinterpretation, one should carefully apply this product in regional-scale studies. One major approach is to improve the quality of the data before application. Several studies have applied the bias-correction method from simple delta change to quantile mapping and those approaches have significantly reduced the errors of model data (Maraun and Widmann, 2018; Teutschbein and Seibert, 2012). Comparison with ERA-20c simulations can be another option for considering the century-term reanalysis in regional climate study. ERA-20c is produced by assimilating surface pressure and marine wind observation, which are not used for ERA-20cm, while both of them are based on the same simulation model, IFS version Cy38r1. For this reason, ERA-20c reanalysis is expected to reproduce higher synoptic realism than ERA-20cm (Poli et al., 2016). Kim and Han (2018) also showed that ERA-20c monthly climate data had better performance than ERA-20cm mean in terms of temporal and statistical agreement to the observed in South Korea.

6. Concluding remarks

In this study, we have evaluated the century-long ERA-20cm ensemble for precipitation and temperature in South Korea. From the evaluations, it could be concluded that ERA-20cm has difficulty in providing useful information on the long term trend as well as the temporal variability for temperature and precipitation in South Korea, although the temperature ensemble has a partial relationship with the observation. For the pdf-based comparison, all ensemble predictions represent significant agreements, but they need improvement for the application in Korea. It is found that the mean may misrepresent ten individual members, especially for the statistical estimate. The ensemble does not spread well enough to cover all observations, especially for temperature, and there is no relationship between the spread and ENSO. This paper also suggests that ERA-20cm variables are biased depending on the region. These findings in this paper help to fill in the knowledge gaps about the applicability of ERA-20cm, one of the latest reanalyses, in South Korea, and provides useful information to readers from other countries on the comparative performance of the global datasets in different parts of the world.

This study has mainly explored the monthly/seasonal/annual mean change on the basis of the averaged dataset over the whole country. Although the analysis is very useful for understanding the general pattern of ERA-20cm ensemble in Korea, it does not represent the extreme climate change and spatial variations. How to improve the results was not deeply dealt with in this study as well. An advantage of ERA-20cm is that they supply the 3-hourly datasets at a higher resolution, 0.125°×0.125°. Hence, the extreme climate change and reduction of the uncertainty will be explored further with the higher spatial and temporal resolutions.

Acknowledgements

The ERA-20cm ensemble data were collected via the ECMWF's public server (http://apps.ecmwf.int/datasets/). Support for the ONI was provided by the NOAA Climate Prediction Centre (http://www.cpc.noaa.gov/). The first author is grateful for the financial support from the Government of South Korea for carrying out his PhD studies at the University of Bristol.

References

Babak, O., Deutsch, C.V., 2009. Statistical approach to inverse distance interpolation. Stoch. Environ. Res. Risk Asses. 23, 543-553.

Bae, D.H., Jung, I.W., Chang, H., 2008. Long-term trend of precipitation and runoff in Korean river basins. Hydrol. Process. 22, 2644-2656.

Bao, X., Zhang, F., 2013. Evaluation of NCEP–CFSR, NCEP–NCAR, ERA-Interim, and ERA-40 reanalysis datasets against independent sounding observations over the Tibetan Plateau. J. Clim. 26, 206-214.

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., Ziese, M., 2013. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901-present. Earth Syst. Sci. Data 5, 71-99.

Befort, D.J., Wild, S., Kruschke, T., Ulbrich, U., Leckebusch, G.C., 2016. Different long-term trends of extra-tropical cyclones and windstorms in ERA-20C and NOAA-20CR reanalyses. Atmospheric Science Letters 17, 586-595.

Bosilovich, M.G., Chen, J., Robertson, F.R., Adler, R.F., 2008. Evaluation of global precipitation in reanalyses. J. Appl. Meteorol. 47, 2279-2299.

Chang, H., Kwon, W.-T., 2007. Spatial variations of summer precipitation trends in South Korea, 1973–2005. Environ. Res. Lett. 2, 045012.

Choi, G., Collins, D., Ren, G., Trewin, B., Baldi, M., Fukuda, Y., Afzaal, M., Pianmana, T., Gomboluudev, P., Huong, P.T.T., 2009. Changes in means and extreme events of temperature and precipitation in the Asia-Pacific Network region, 1955–2007. Int. J. Climatol. 29, 1906-1925.

Choi, K.-S., Kim, B.-J., Zhang, R., Nam, J.-C., Park, K.-J., Kim, J.-Y., Kim, D.-W., 2016. Possible influence of South Asian high on summer rainfall variability in Korea. Clim. Dynam. 46, 833-846.

Chung, U., Choi, J., Yun, J.I., 2004. Urbanization effect on the observed change in mean monthly temperatures between 1951-1980 and 1971-2000 in Korea. Clim. Change 66, 127-136.

Chung, Y., Yoon, M., 2000. Interpretation of recent temperature and precipitationtrends observed in Korea. Theor. Appl. Climatol. 67, 171-180.

Compo, G.P., Whitaker, J.S., Sardeshmukh, P.D., Matsui, N., Allan, R.J., Yin, X., Gleason, B.E., Vose, R.S., Rutledge, G., Bessemoulin, P., 2011. The twentieth century reanalysis project. Q. J. R. Meteorol. Soc. 137, 1-28.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.,

Balsamo, G., Bauer, P., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553-597.
Donat, M.G., Alexander, L.V., Herold, N., Dittus, A.J., 2016. Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. J. Geophys. Res.-Atmos. 121, 11,174-111,189.
Donat, M.G., Sillmann, J., Wild, S., Alexander, L.V., Lippmann, T., Zwiers, F.W., 2014. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. J. Clim. 27, 5019-5035.
Ferguson, C.R., Villarini, G., 2012. Detecting inhomogeneities in the Twentieth Century Reanalysis over the central United States. J. Geophys. Res.-Atmos. 117.
Gao, L., Bernhardt, M., Schulz, K., 2012. Elevation correction of ERA-Interim temperature data in complex terrain. Hydrology and Earth System Sciences 16, 4661-4673.
Gao, L., Bernhardt, M., Schulz, K., Chen, X.W., Chen, Y., Liu, M.B., 2016. A First Evaluation of ERA-20CM over China. Mon. Weather Rev. 144, 45-57.
Gao, L., Hao, L., Chen, X.-w., 2014a. Evaluation of ERA-interim monthly temperature data over the Tibetan Plateau. Journal of Mountain Science 11, 1154-1168.
Gao, L., Schulz, K., Bernhardt, M., 2014b. Statistical downscaling of ERA-interim forecast precipitation data in complex terrain using lasso algorithm. Advances in Meteorology 2014, 472741.
Hamed, K.H., Rao, A.R., 1998. A modified Mann-Kendall trend test for autocorrelated data. J. Hydrol. 204, 182-196.
Harris, I., Jones, P.D., Osborn, T.J., Lister, D.H., 2014. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. Int. J. Climatol. 34, 623-642.
Hersbach, H., Peubey, C., Simmons, A., Berrisford, P., Poli, P., Dee, D., 2015. ERA-20CM: a twentieth-century atmospheric model ensemble. Q. J. R. Meteorol. Soc. 141, 2350-2375.
Ho, C.H., Lee, J.Y., Ahn, M.H., Lee, H.S., 2003. A sudden change in summer rainfall characteristics in Korea during the late 1970s. Int. J. Climatol. 23, 117-128.
Jeong, M.S., Kim, J.S., Oh, S.M., Moon, Y.I., 2015. Sensitivity of summer precipitation over the Korean Peninsula to temperature gradients. Int. J. Climatol. 35, 836-845.
Jin, Y.H., Kawamura, A., Jinno, K., Berndtsson, R., 2005. Detection of ENSO-influence on the monthly precipitation in South Korea. Hydrol. Process. 19, 4081-4092.
Jung, H.S., Choi, Y., Oh, J.H., Lim, G.H., 2002. Recent trends in temperature and precipitation over South Korea. Int. J. Climatol. 22, 1327-1337.
Jung, I.W., Bae, D.H., Kim, G., 2011. Recent trends of mean and extreme precipitation in Korea. Int.

J. Climatol. 31, 359-370.
Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77, 437-471.
Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., Potter, G., 2002. Ncep–doe amip-ii reanalysis (r-2). Bull. Am. Meteorol. Soc. 83, 1631-1643.
Kendall, M.G. 1955. Rank correlation methods, Griffin, London.
Kim, D.-I., Han, D., 2018. Comparative study on long term climate data sources over South Korea. Journal of Water and Climate Change (In press).
Krueger, O., Schenk, F., Feser, F., Weisse, R., 2013. Inconsistencies between long-term trends in storminess derived from the 20CR reanalysis and observations. J. Clim. 26, 868-874.
Lee, J.H., Julien, P.Y., 2016. ENSO impacts on temperature over South Korea. Int. J. Climatol. 36, 3651-3663.
Li, L., Xia, J., Xu, C.-Y., Singh, V., 2010. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. J. Hydrol. 390, 210-221.
Ma, L., Zhang, T., Frauenfeld, O.W., Ye, B., Yang, D., Qin, D., 2009. Evaluation of precipitation from the ERA-40, NCEP-1, and NCEP-2 Reanalyses and CMAP-1, CMAP-2, and GPCP-2 with ground-based measurements in China. J. Geophys. Res.-Atmos. 114, D09105.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13, 245-259.
Maraun, D., 2013. Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. J. Clim. 26, 2137-2143.
Maraun, D., Widmann, M. 2018. Statistical Downscaling and Bias Correction for Climate Research, Cambridge University Press.
Perkins, S.E., Pitman, A.J., Holbrook, N.J., McAneney, J., 2007. Evaluation of the AR4 climate models' simulated daily maximum temperature, minimum temperature, and precipitation over Australia using probability density functions. J. Clim. 20, 4356-4376.
Poli, P., Hersbach, H., Dee, D.P., Berrisford, P., Simmons, A.J., Vitart, F., Laloyaux, P., Tan, D.G., Peubey, C., Thépaut, J.-N., 2016. ERA-20C: An Atmospheric Reanalysis of the Twentieth Century. J. Clim. 29, 4083-4097.
Poli, P., Hersbach, H., Tan, D., Dee, D., Thepaut, J.-N., Simmons, A., Peubey, C., Laloyaux, P., Komori, T., Berrisford, P. 2013. The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C).
Sayemuzzaman, M., Jha, M.K., 2014. Seasonal and annual precipitation time series trend analysis in

North Carolina, United States. Atmos. Res. 137, 183-194.
Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. 63, 1379-1389.
Shadmani, M., Marofi, S., Roknian, M., 2012. Trend analysis in reference evapotranspiration using Mann-Kendall and Spearman's Rho tests in arid regions of Iran. Water Resour. Manage. 26, 211-224.
Shah, R., Mishra, V., 2014. Evaluation of the reanalysis products for the monsoon season droughts in India. Journal of Hydrometeorology 15, 1575-1591.
Simmons, A.J., Jones, P.D., Bechtold, V.D., Beljaars, A.C.M., Kallberg, P.W., Saarinen, S., Uppala, S.M., Viterbo, P., Wedi, N., 2004. Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. J. Geophys. Res.-Atmos. 109, D24115.
Son, H.-Y., Park, J.-Y., Kug, J.-S., Yoo, J., Kim, C.-H., 2014. Winter precipitation variability over Korean Peninsula associated with ENSO. Clim. Dynam. 42, 3171-3186.
Teutschbein, C., Seibert, J., 2012. Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods. J. Hydrol. 456, 12-29.
Theil, H., 1950. A rank-invariant method of linear and polynomial regression analysis. I, II, III. Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen 53, 386–392, 521–525, 1397–1412.
Thorne, P., Vose, R., 2010. Reanalyses suitable for characterizing long-term trends: are they really achievable? Bull. Am. Meteorol. Soc. 91, 353-361.
Uppala, S.M., Kållberg, P., Simmons, A., Andrae, U., Bechtold, V.d., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., 2005. The ERA-40 re-analysis. Q. J. R. Meteorol. Soc. 131, 2961-3012.
Wang, B., Ding, Q., Jhun, J.G., 2006. Trends in Seoul (1778–2004) summer precipitation. Geophys. Res. Lett. 33, L15803.
Xu, Z., Takeuchi, K., Ishidaira, H., Li, J., 2005. Long-term trend analysis for precipitation in Asian Pacific FRIEND river basins. Hydrol. Process. 19, 3517-3532.
Zhao, T., Fu, C., 2006. Comparison of products from ERA-40, NCEP-2, and CRU with station data for summer precipitation over China. Advances in Atmospheric sciences 23, 593-604.

Tables

Table 1 The local weather stations shown in Fig. 1

No.	Name	Longitude	Latitude	Observation Period	Elevation (m. asl)
1	Seoul	126-57-56 E	37-34-17 N	1961-2010	11.1
2	Incheon	126-37-29 E	37-28-39 N	1961-2010	69.6
3	Seosan	126-29-45 E	36-46-25 N	1968-2010	30.3
4	Chuncheon	127-44-08 E	37-54-09 N	1966-2010	79.1
5	Gangneung	128-53-27 E	37-45-05 N	1961-2010	27.4
6	Jeonju	127-09-17 E	35-49-17 N	1961-2010	54.8
7	Chupungnyeong	127-59-40 E	36-13-11 N	1961-2010	246.1
8	Yeongju	128-31-00 E	36-52-18 N	1973-2010	212.2
9	Gwangju	126-53-29 E	35-10-22 N	1961-2010	73.8
10	Yeosu	127-44-26 E	34-44-21 N	1961-2010	66.0
11	Daegu	128-37-08 E	35-53-06 N	1961-2010	65.5
12	Pohang	129-22-46 E	36-01-57 N	1961-2010	3.7
13	Busan	129-01-55 E	35-06-16 N	1961-2010	71.0

Table 2 Correlation coefficient(r) for seasonal and annual total precipitation between the observation and each dataset averaged over all regions from 1961 to 2010

Type	Seasonal				Annual
	spring	summer	autumn	winter	
Mean	-0.009	0.077	-0.035	0.278	0.103
En0	0.065	0.207	-0.146	0.184	0.147
En1	-0.041	-0.087	0.028	0.137	-0.065
En2	-0.166	0.081	-0.008	0.170	0.069
En3	0.045	-0.001	0.030	-0.029	0.074
En4	-0.045	0.038	0.054	0.173	-0.027
En5	0.005	-0.057	-0.101	0.189	-0.013
En6	0.128	0.102	-0.061	0.130	0.158
En7	-0.171	0.053	0.010	0.045	0.059
En8	0.003	-0.081	0.020	0.228	-0.046
En9	0.133	0.062	0.026	0.227	0.116

Table 3 Correlation coefficient(r) for annual total precipitation between ensemble members from 1961 to 2010

	En0	En1	En2	En3	En4	En5	En6	En7	En8	En9
En0	-	0.173	0.285	0.125	0.199	0.101	0.338	0.160	0.076	0.091
En1	0.173	-	0.270	-0.008	0.305	0.095	0.091	-0.079	0.068	0.108
En2	0.285	0.270	-	0.179	0.048	0.112	0.193	-0.113	0.058	0.304
En3	0.125	-0.008	0.179	-	0.222	0.101	0.325	0.206	-0.059	0.111
En4	0.199	0.305	0.048	0.222	-	0.144	0.105	0.286	0.088	0.064
En5	0.101	0.095	0.112	0.101	0.144	-	0.274	0.212	0.021	0.094
En6	0.338	0.091	0.193	0.325	0.105	0.274	-	0.114	0.163	0.083
En7	0.160	-0.079	-0.113	0.206	0.286	0.212	0.114	-	-0.009	0.311
En8	0.076	0.068	0.058	-0.059	0.088	0.021	0.163	-0.009	-	0.030
En9	0.091	0.108	0.304	0.111	0.064	0.094	0.083	0.311	0.030	-

Table 4 Mann-Kendall test results for precipitation trend

Dataset	Spring		Summer		Autumn		Winter		Annual	
	Z	β	Z	β	Z	β	Z	β	Z	β
1961-2010										
Observation	-0.23	-0.34	2.19 [a]	3.46	-0.74	-0.73	0.28	0.09	0.82	2.25
Mean	0.70	0.16	0.43	0.31	1.14	0.44	0.20	0.07	1.36	1.35
En0	0.52	0.40	0.33	0.61	1.30	0.96	0.64	0.46	0.87	1.67
En1	0.92	0.62	-0.65	-0.93	0.69	0.72	1.37	0.69	0.82	1.43
En2	0.80	0.70	0.33	0.57	1.69 [b]	1.33	1.57	0.97	1.39	2.78
En3	1.30	1.24	-0.79	-1.73	-0.33	-0.28	-0.40	-0.29	-0.18	-0.56
En4	0.03	0.04	0.40	0.46	1.47	1.18	-0.13	-0.06	0.80	1.60
En5	0.95	0.72	0.99	1.36	0.42	0.30	-0.05	-0.03	1.29	2.49
En6	-0.94	-0.76	2.16 [a]	3.14	0.72	0.64	0.00	-0.01	1.05	2.17
En7	-0.62	-0.51	-0.05	-0.09	-0.07	-0.07	-1.05	-0.69	-0.32	-0.60
En8	-0.79	-0.57	-0.40	-0.66	-0.17	-0.11	0.38	0.21	-0.55	-1.12
En9	0.37	0.32	0.59	1.06	-0.67	-0.51	0.00	0.00	1.36	2.34
1901-2010										
Mean	2.58 [a]	0.20	0.52	0.11	1.19	0.12	1.82 [b]	0.15	2.14 [a]	0.60
En0	0.11	0.02	0.80	0.38	-0.60	-0.13	0.24	0.04	0.30	0.18
En1	-0.68	-0.15	-0.07	-0.02	-0.54	-0.13	0.70	0.10	0.01	0.01
En2	0.74	0.17	-0.94	-0.48	2.19 [a]	0.51	0.50	0.09	0.65	0.37
En3	2.58 [a]	0.57	-1.90 [b]	-0.89	-0.23	-0.06	1.42	0.25	-0.12	-0.06
En4	1.79 [b]	0.39	-0.85	-0.39	1.83 [b]	0.43	1.01	0.16	0.95	0.58
En5	2.36 [a]	0.55	0.81	0.41	-0.56	-0.12	0.37	0.05	1.20	0.85
En6	0.24	0.04	1.44	0.70	1.96 [a]	0.45	0.26	0.05	1.55	0.97
En7	1.25	0.27	-0.09	-0.05	0.77	0.19	2.09 [a]	0.36	1.30	0.79
En8	1.32	0.30	-0.01	-0.01	-0.84	-0.16	1.29	0.19	0.79	0.52
En9	0.22	0.06	2.04 [a]	1.00	0.09	0.02	0.50	0.08	2.57 [a]	1.39

Z : the standardized test statistic in M-K test

[a] : significant trend at the 0.05 significance level. [b] : significant trend at the 0.10 significance level.

β(trends for precipitation) are in mm/yr

Table 5 The percent of the observation for precipitation bracketed by the ERA-20cm ensemble intervals

Type	Interannual					PDF
	Spring	Summer	Autumn	Winter	Annual	
P(%)	66.0	56.0	74.0	60.0	70.0	42.3 (34.6*)

* : The percent excluding the number of observation on-boundary

Table 6 Correlation coefficient(r) for seasonal and annual mean temperature between the observation and each dataset averaged over all regions from 1961 to 2010

Type	Seasonal				Annual
	spring	summer	autumn	winter	
Mean	0.691	0.573	0.612	0.471	0.746
En0	0.509	0.265	0.319	0.244	0.592
En1	0.507	0.394	0.241	0.076	0.521
En2	0.418	0.480	0.409	0.414	0.547
En3	0.356	0.452	0.404	0.048	0.443
En4	0.307	0.360	0.482	0.242	0.526
En5	0.450	0.356	0.426	0.339	0.561
En6	0.373	0.525	0.411	0.193	0.586
En7	0.515	0.364	0.428	0.278	0.580
En8	0.341	0.475	0.391	0.324	0.546
En9	0.403	0.443	0.392	0.350	0.604

Table 7 Correlation coefficient(r) for annual mean temperature between ensemble members from 1961 to 2010

	En0	En1	En2	En3	En4	En5	En6	En7	En8	En9
En0	-	0.572	0.481	0.568	0.530	0.490	0.602	0.616	0.511	0.611
En1	0.572	-	0.471	0.450	0.512	0.327	0.495	0.438	0.565	0.526
En2	0.481	0.471	-	0.422	0.268	0.441	0.390	0.455	0.361	0.523
En3	0.568	0.450	0.422	-	0.635	0.355	0.568	0.445	0.421	0.628
En4	0.530	0.512	0.268	0.635	-	0.297	0.491	0.421	0.385	0.593
En5	0.490	0.327	0.441	0.355	0.297	-	0.462	0.472	0.300	0.538
En6	0.602	0.495	0.390	0.568	0.491	0.462	-	0.566	0.481	0.682
En7	0.616	0.438	0.455	0.445	0.421	0.472	0.566	-	0.494	0.608
En8	0.511	0.565	0.361	0.421	0.385	0.300	0.481	0.494	-	0.375
En9	0.611	0.526	0.523	0.628	0.593	0.538	0.682	0.608	0.375	-

Table 8 Mann-Kendall test results for temperature trend

Dataset	Spring		Summer		Autumn		Winter		Annual	
	Z	β	Z	β	Z	β	Z	β	Z	β
1961-2010										
Observation	3.45 [a]	2.54	0.89	0.59	2.24 [a]	1.82	3.09 [a]	3.63	3.66 [a]	2.09
Mean	3.85 [a]	1.44	4.00 [a]	1.72	2.79 [a]	1.26	1.59	0.54	3.75 [a]	1.34
En0	1.74 [b]	0.89	2.31 [a]	1.18	1.30	1.02	1.84 [b]	1.00	2.56 [a]	1.16
En1	2.91 [a]	1.72	3.58 [a]	1.94	1.82 [b]	1.31	-1.12	-0.88	1.97 [a]	1.13
En2	2.19 [a]	1.44	2.12 [a]	1.22	1.52	0.90	0.77	0.53	2.78 [a]	1.11
En3	2.33 [a]	1.71	1.22	0.95	2.61 [a]	1.68	0.72	0.45	2.64 [a]	1.31
En4	2.53 [a]	1.18	2.88 [a]	1.99	1.87 [b]	1.20	2.12 [a]	1.56	3.73 [a]	1.56
En5	1.49	0.97	2.94 [a]	1.17	1.89 [b]	1.31	0.65	0.43	2.66 [a]	1.12
En6	1.91 [b]	1.53	4.28 [a]	2.41	2.99 [a]	1.68	-0.28	-0.10	3.06 [a]	1.36
En7	2.23 [a]	1.47	2.43 [a]	1.43	1.42	1.07	0.38	0.35	2.29 [a]	1.15
En8	2.28 [a]	1.31	2.84 [a]	1.49	0.62	0.33	0.67	0.50	2.88 [a]	1.11
En9	3.28 [a]	2.05	3.45 [a]	2.35	2.96 [a]	1.96	1.25	0.94	3.93 [a]	2.04
1901-2010										
Mean	8.69 [a]	1.02	7.60 [a]	0.91	6.61 [a]	0.81	6.61 [a]	0.79	9.38 [a]	0.95
En0	4.90 [a]	1.02	4.02 [a]	0.60	4.17 [a]	0.80	3.19 [a]	0.61	6.33 [a]	0.80
En1	4.18 [a]	0.79	6.45 [a]	1.13	4.25 [a]	0.87	3.59 [a]	0.81	6.89 [a]	0.99
En2	5.48 [a]	1.11	5.75 [a]	0.98	4.23 [a]	0.72	4.84 [a]	1.05	7.80 [a]	1.03
En3	4.86 [a]	0.93	4.52 [a]	0.85	3.49 [a]	0.66	2.84 [a]	0.61	6.10 [a]	0.84
En4	4.42 [a]	0.67	5.80 [a]	1.16	4.48 [a]	0.85	4.57 [a]	0.92	7.17 [a]	0.96
En5	6.05 [a]	1.24	4.45 [a]	0.70	4.45 [a]	0.89	4.19 [a]	0.87	7.38 [a]	1.02
En6	2.85 [a]	0.62	5.16 [a]	0.95	3.72 [a]	0.64	2.06 [a]	0.44	5.33 [a]	0.70
En7	5.66 [a]	1.08	5.99 [a]	1.09	3.14 [a]	0.64	4.09 [a]	0.87	7.00 [a]	1.04
En8	6.05 [a]	1.27	5.06 [a]	0.85	4.25 [a]	0.93	2.96 [a]	0.70	6.77 [a]	0.99
En9	6.48 [a]	1.39	4.91 [a]	0.88	4.74 [a]	0.99	3.75 [a]	0.80	7.29 [a]	1.13

Z : the standardized test statistic in M-K test

[a] : significant trend at the 0.05 significance level. [b] : significant trend at the 0.10 significance level.

β(trends for temperature) are in $10^{-2\circ}\ C/yr$

Table 9 The percent of the observation for temperature bracketed by the ERA-20cm ensemble intervals

Type	Interannual					PDF
	Spring	Summer	Autumn	Winter	Annual	
P(%)	4.0	28.0	22.0	42.0	4.0	37.1 (28.6*)

*: The percent excluding the number of observation on-boundary

Figures

Fig. 1. Locations of 13 gauge stations and gridded points of ERA-20cm. The grey shading on the map indicates elevations.

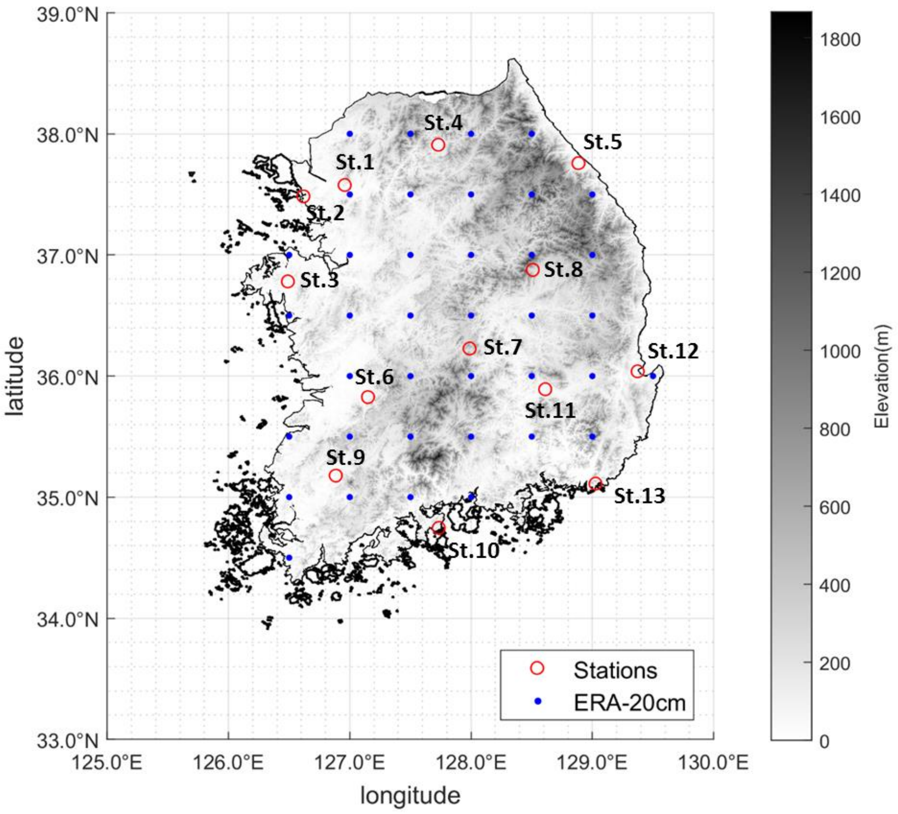
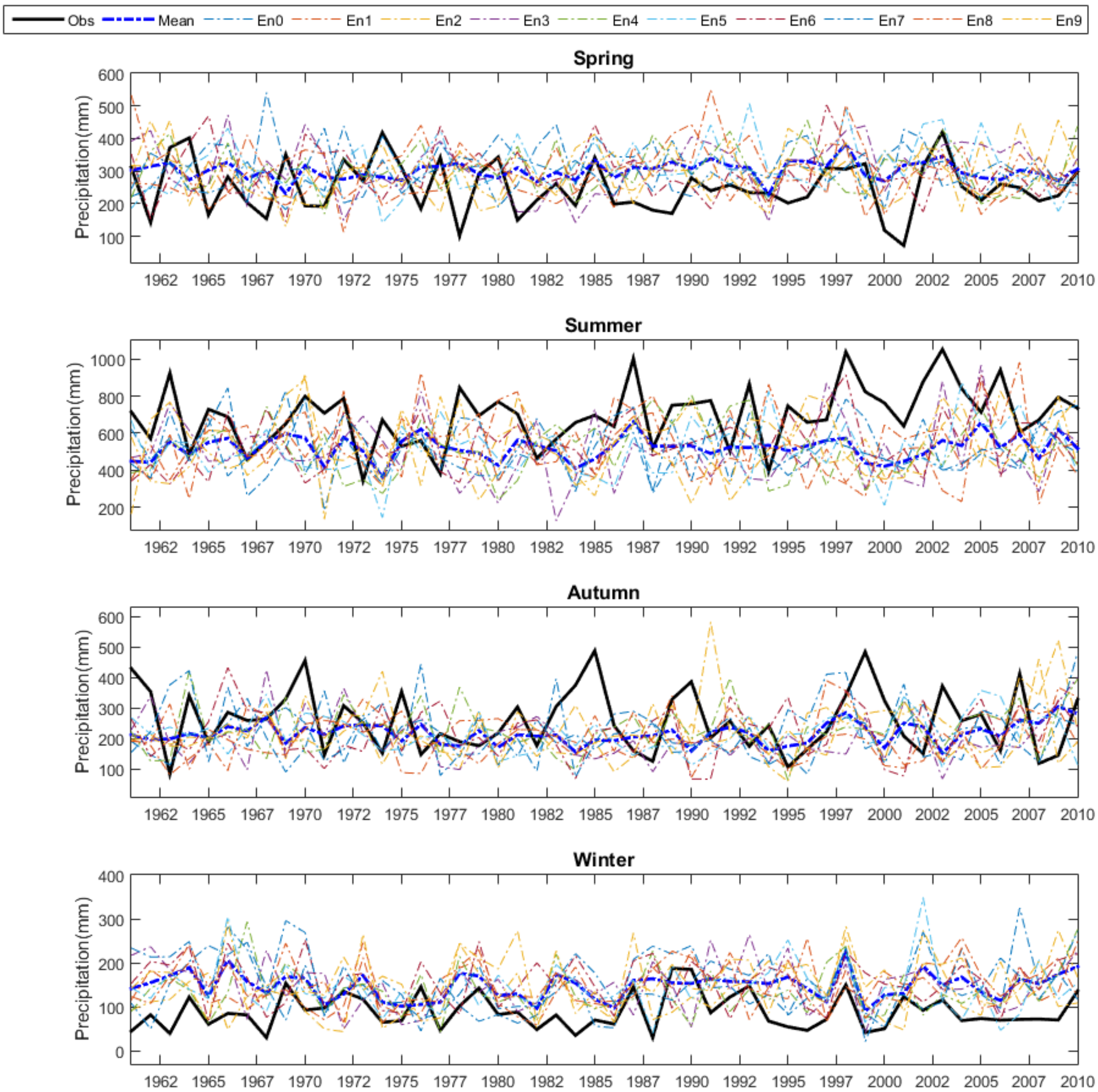


Fig. 2. Total precipitation change averaged over the whole region from 1961 to 2010

(a) The seasonal total precipitation change (from top to bottom, Spring, Summer, Autumn, and Winter)



(b) The annual total precipitation change

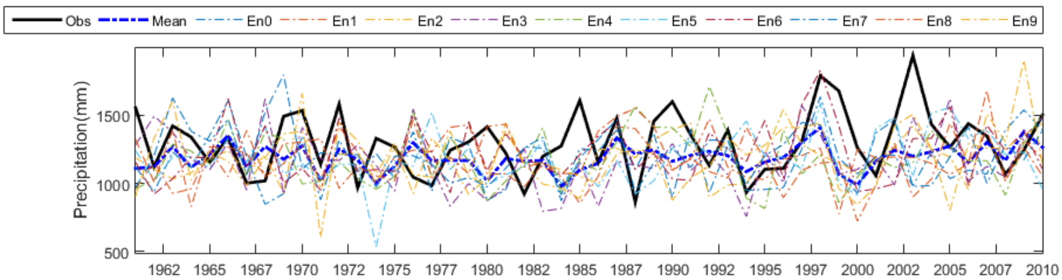


Fig. 3. Relationship between 3month moving average of ERA-20cm ensemble variance for precipitation in South Korea (top) and El Niños/La Niñas events (bottom) from 1961 to 2010

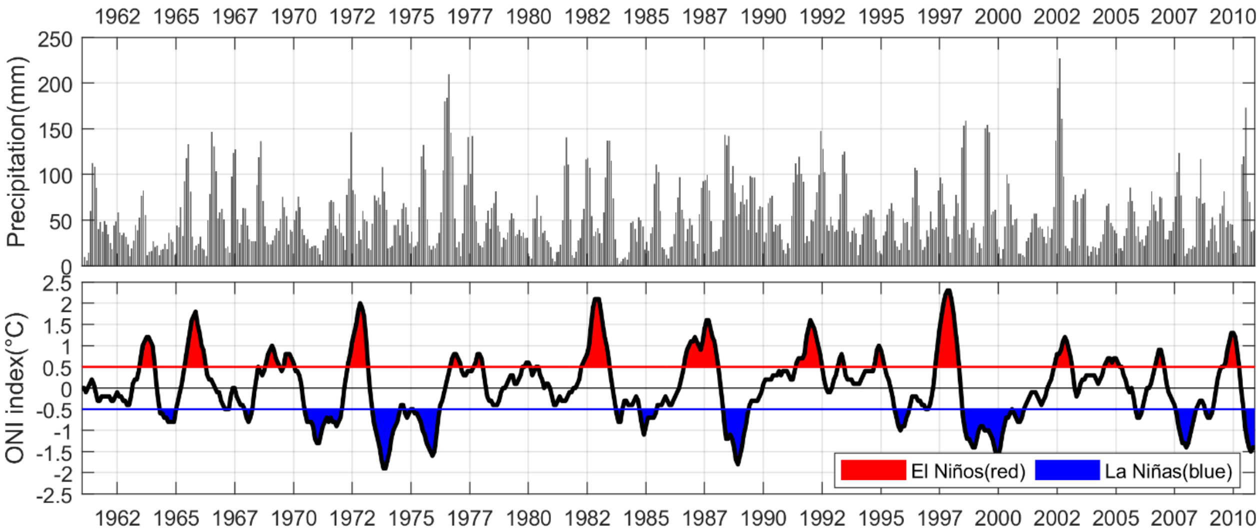


Fig. 4. PDF-based skill score for monthly precipitation for each dataset averaged over the whole region from 1961 to 2010

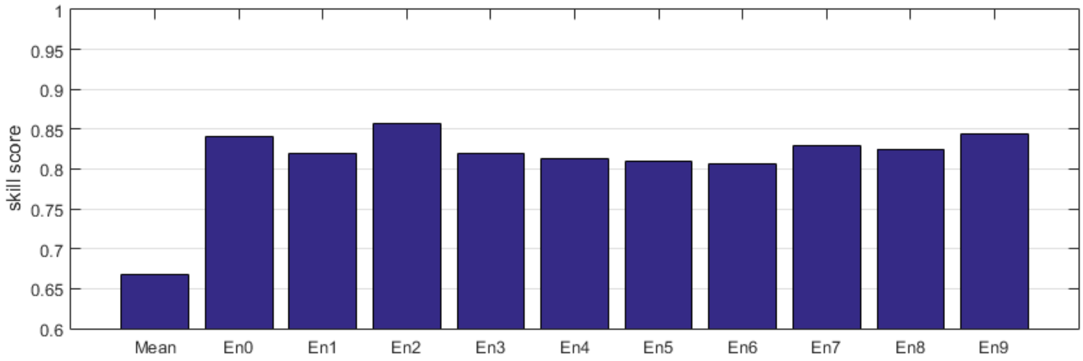
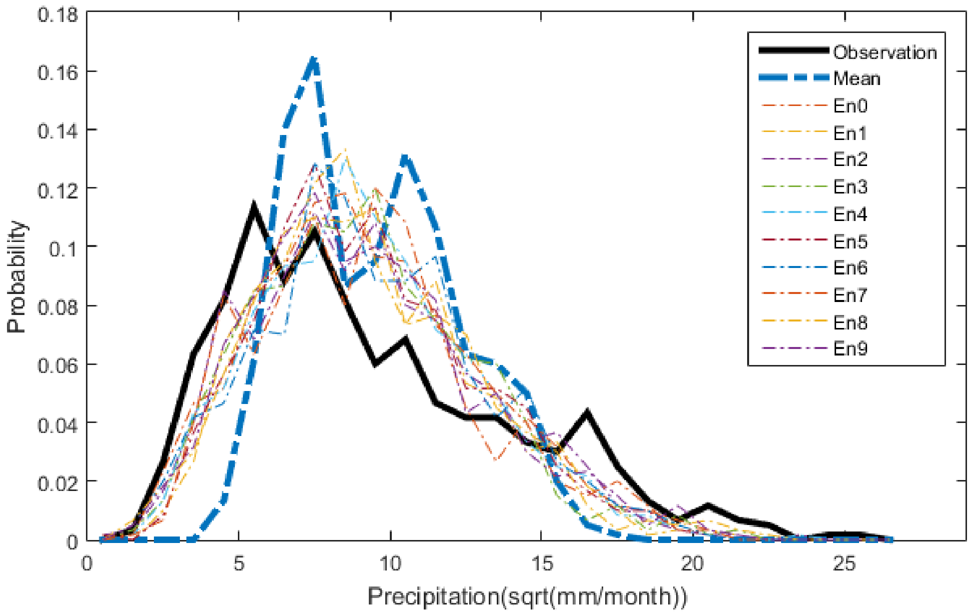


Fig. 5. Probability density functions (PDFs) for monthly total precipitation over South Korea

(a) PDFs for monthly total precipitation from 1961 to 2010



(b) PDFs for seasonally subdivided monthly total precipitation from 1961 to 2010

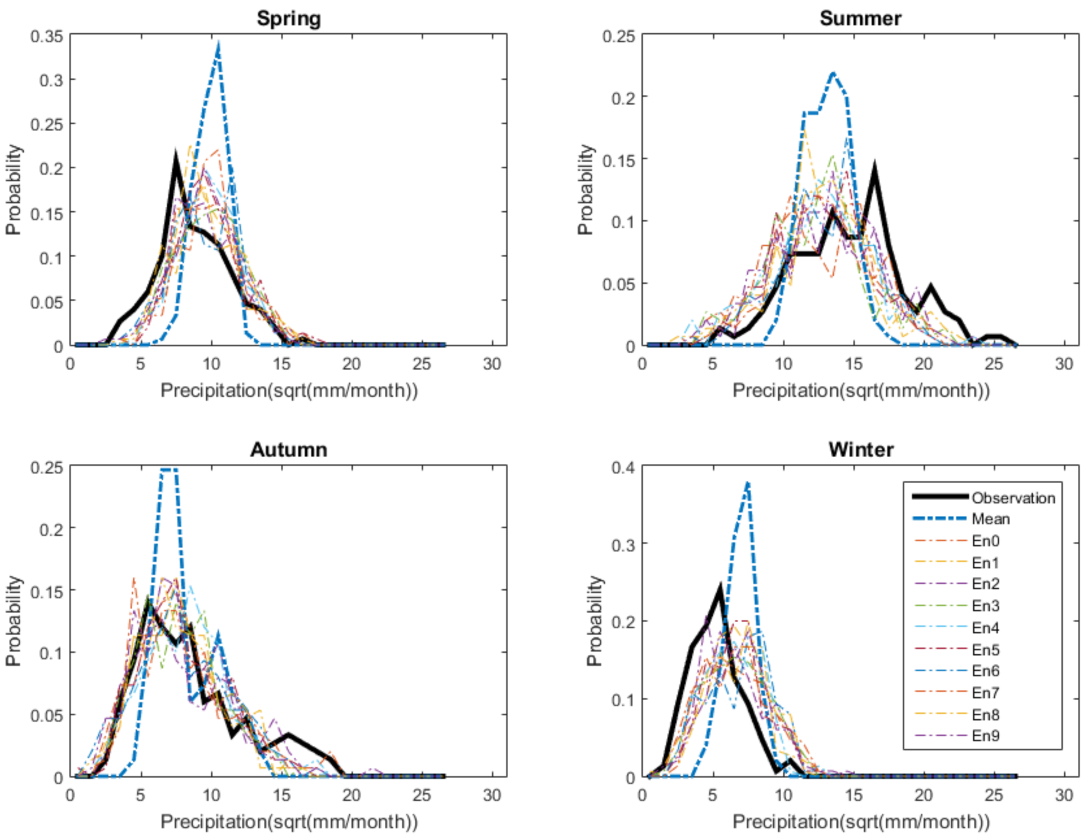
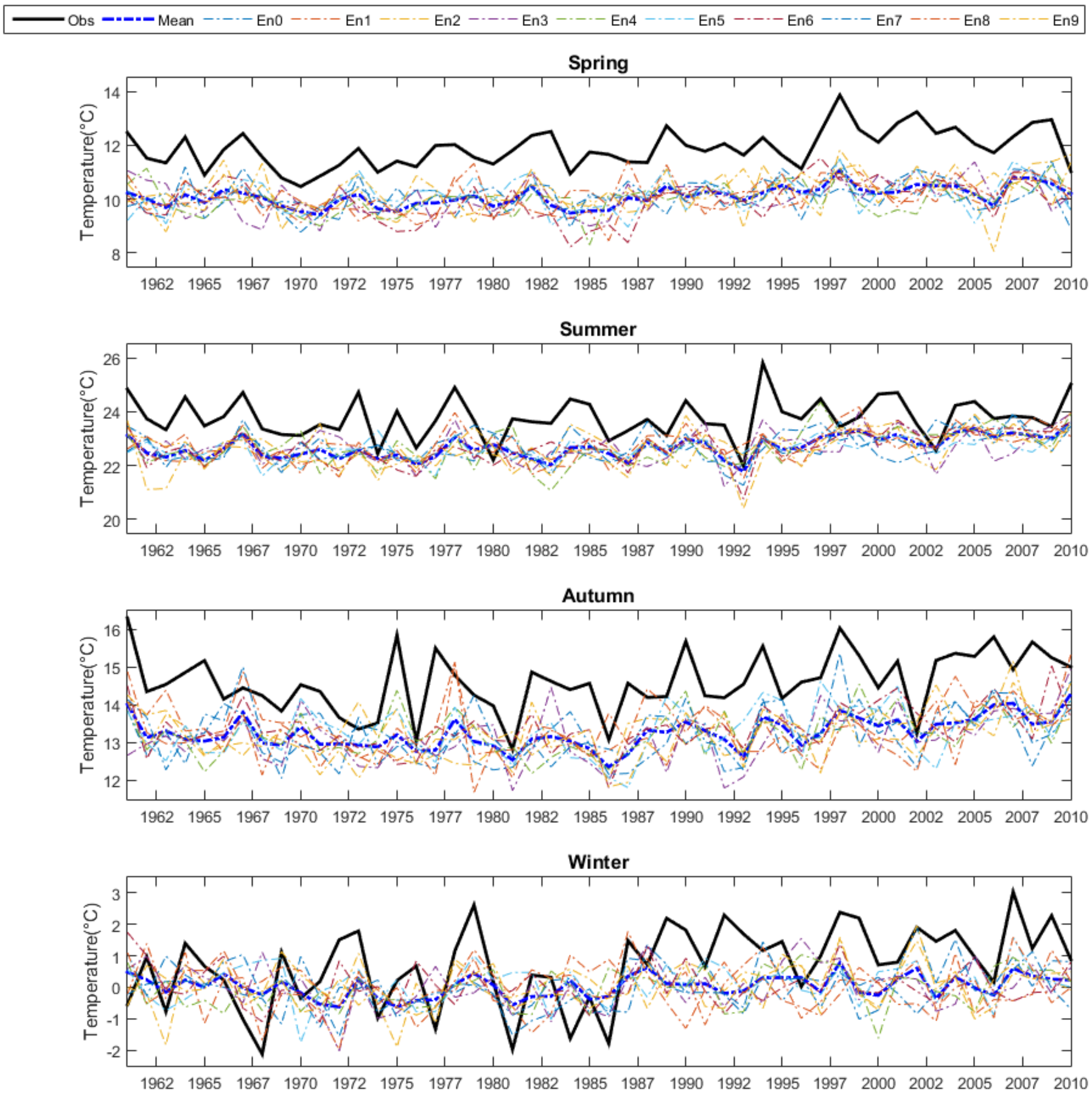


Fig. 6. Mean temperature change averaged over the whole region from 1961 to 2010

(a) The seasonal mean temperature change (from top to bottom, Spring, Summer, Autumn, and Winter)



(b) The annual mean temperature change

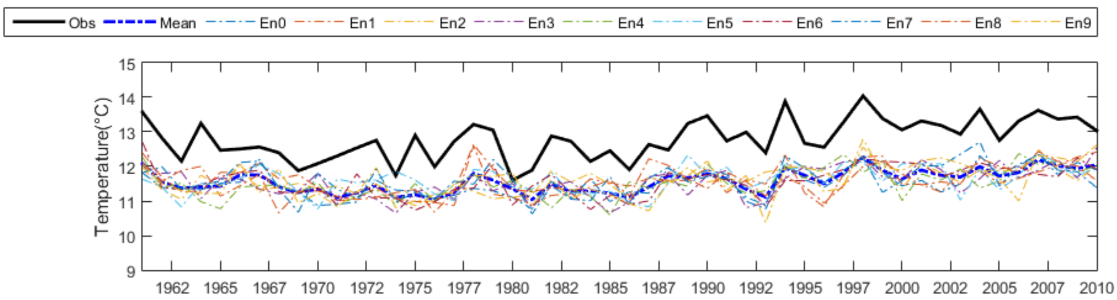


Fig. 7. Relationship between 3month moving average of ERA-20cm ensemble variance for temperature in South Korea (top) and El Niños/La Niñas events (bottom) from 1961 to 2010

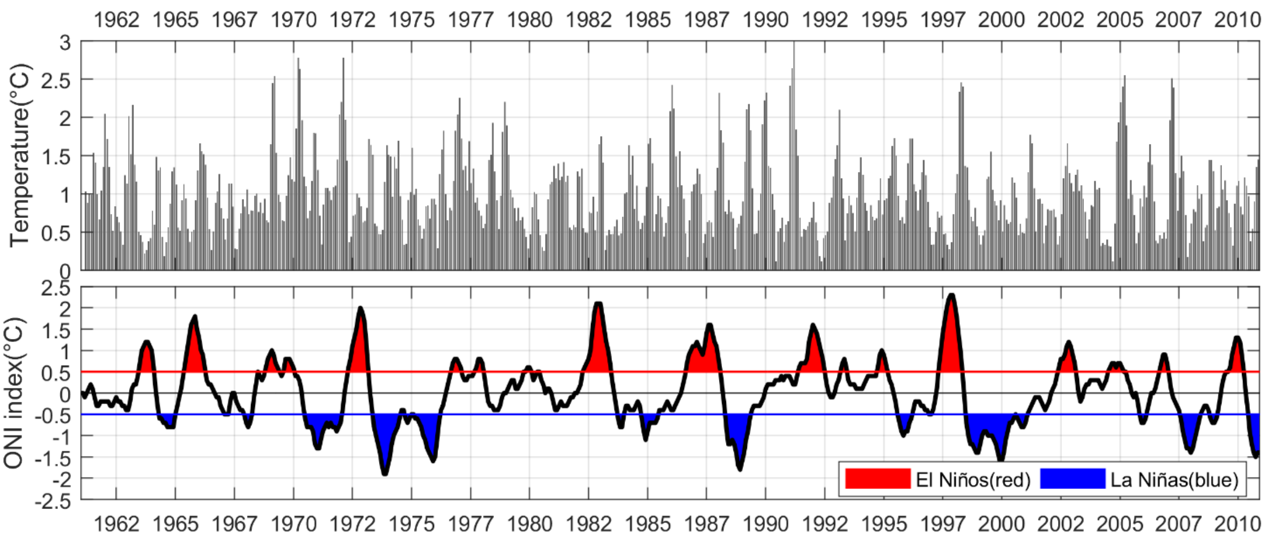


Fig. 8. PDF-based skill score for monthly mean temperature for each dataset averaged over the whole region from 1961 to 2010

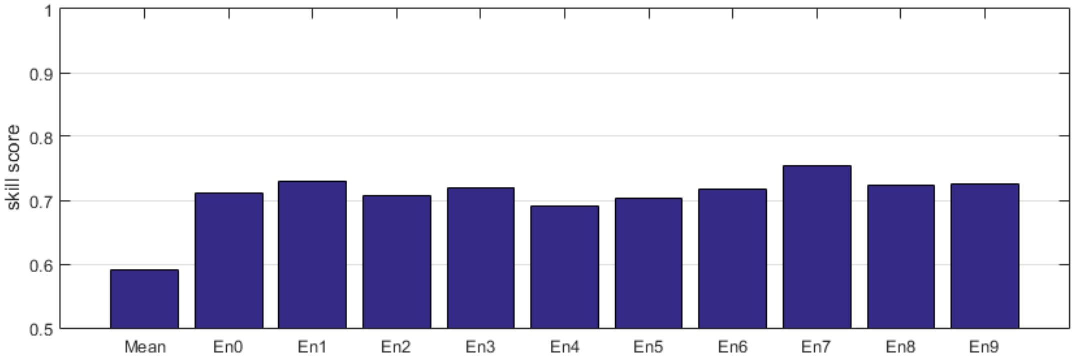
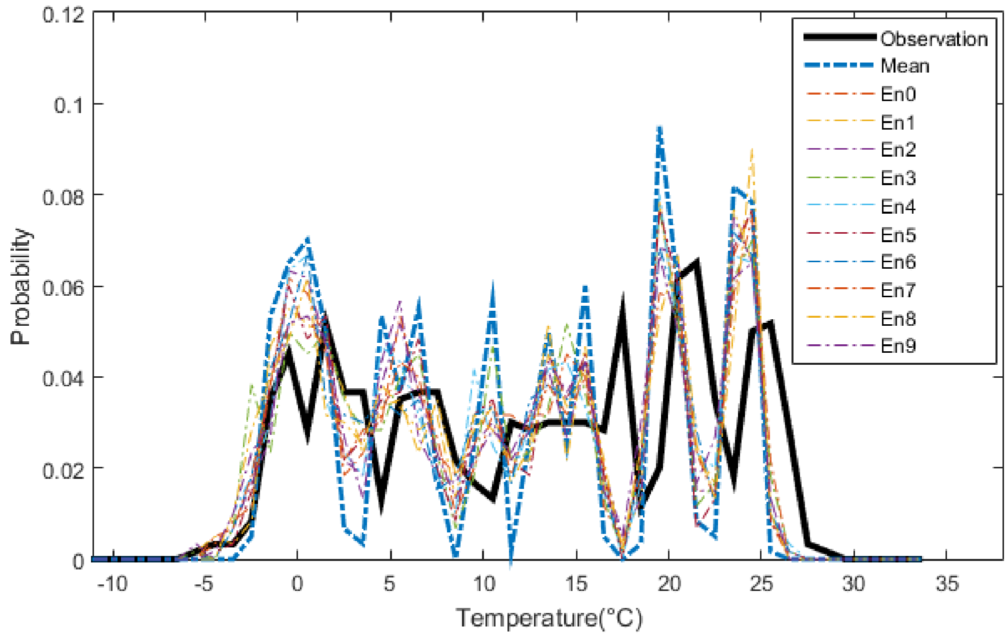


Fig. 9. Probability density functions for monthly mean temperature over South Korea

(a) PDFs for monthly mean temperature from 1961 to 2010



(b) PDFs for seasonally subdivided monthly mean temperature from 1961 to 2010